

金融数据库

■ 朱世武 严玉星 编著

清华大学出版社



F830.49/30

2007

金融数据库

朱世武 严玉星 编著

清华大学出版社
北 京

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13501256678 13801310933

图书在版编目(CIP)数据

金融数据库/朱世武,严玉星编著. —北京:清华大学出版社,2007.12

ISBN 978-7-302-16535-4

I. 金… II. ①朱…②严… III. 金融—数据库 IV. F83-39

中国版本图书馆 CIP 数据核字(2007)第 184256 号

责任编辑:江 娅

责任校对:王凤芝

责任印制:王秀菊

出版发行:清华大学出版社 地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn> 邮 编:100084

c-service@tup.tsinghua.edu.cn

社 总 机:010-62770175 邮购热线:010-62786544

投稿咨询:010-62772015 客户服务:010-62776969

印 装 者:北京鑫海金澳胶印有限公司

经 销:全国新华书店

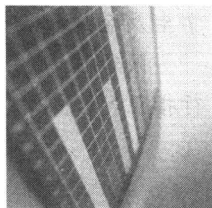
开 本:185×260 印 张:17.25 字 数:404 千字

版 次:2007 年 12 月第 1 版 印 次:2007 年 12 月第 1 次印刷

印 数:1~4000

定 价:28.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770177 转 3103 产品编号:027112-01



前言

金融数据库 · PREFACE

现代经济、金融学等学科的发展越来越需要数据平台的支持。离开数据,根本就不可能作实证研究。近年来,国内外各大高等院校纷纷开设金融数据库、金融计算与建模等相关课程。目前,已有不少金融计算与建模的教材相继出版,但国内还没有一本金融数据库的教材。本书作者在多年讲授国内外金融数据库和提供数据服务技术支持的经验基础上,完成了本书的写作。本书的出版填补了国内金融数据库教材的空白,希望本书的出版能对金融数据库、金融计算与建模学科的发展起到一定的促进作用。

全书共分10章。第1章为金融数据库概述,介绍国内外著名金融数据库,并阐述了金融数据库的作用、分类及相关质量评价标准。第2~6章以 RESSET 金融研究数据库 (RESSET/DB) 为平台,介绍中国股票市场、上市公司财务、基金、固定收益、宏观与行业的数据结构、内容与相关数据的获取方式。第7~10章介绍国际著名数据库 CRSP, Compustat, IBES 和 NYSE TAQ 的数据内容与相关获取方式。

通过阅读本书,读者可以了解相关数据库的内容及使用方法。本书附录可帮助读者查询与组织数据,并使读者具备利用 SAS 等软件进行复杂数据处理的能力。

本书得到了专业金融数据网站(www.resset.cn)的在线技术支持,提供配套数据库、程序下载与疑难问题解答等服务,以方便读者学习。

本书既可以帮助读者从数据层面出发了解国内外金融市场的结构、交易机制等,也是综合的金融数据库使用手册。

本书适合多层次专业人员阅读及使用,如金融经济、数学和统计学等专业的本科生、研究生及相关研究人员。

本书1~6章由朱世武完成。7~10章由严玉星完成英文稿,朱世武、于佳、严蓉鸽、冯雪和叶欢译成中文,并由严玉星、朱世武对中文稿进行了全文审定。

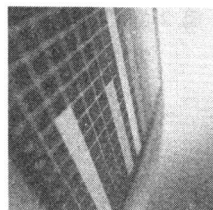
本书写作过程中,得到了锐思数据(www.resset.cn)的大力帮助。特别感谢锐思数据的王时雨、陶中伟、于佳与陈燕飞,感谢清华学生张小红、赵宏旭的帮助。同时也感谢其他高校

学生及金融机构专业人士提出的宝贵建议。

限于作者的水平,书中难免存在不少不足之处,敬请读者提出宝贵建议并对错误之处进行指正,便于以后再版时加以改进。

作 者

2007年10月于清华园



目 录

金融数据库 · CONTENTS

第 1 章 金融数据库概论	(1)
1.1 金融数据库的起源	(1)
1.2 金融数据库的作用	(2)
1.3 金融数据库的类别	(2)
1.4 中外金融数据库概况	(3)
1.4.1 CRSP	(4)
1.4.2 Compustat	(5)
1.4.3 PACAP	(6)
1.4.4 Datastream	(6)
1.4.5 NYSE TAQ	(7)
1.4.6 IBES	(8)
1.4.7 SDC	(8)
1.4.8 GovPX	(9)
1.4.9 Reuter	(10)
1.4.10 Bloomberg	(10)
1.4.11 聚源	(10)
1.4.12 世华	(11)
1.4.13 万得	(11)
1.4.14 锐思数据	(11)
1.4.15 国泰安	(12)
1.5 金融数据库的选择标准	(12)
1.6 选择合适的数据库	(14)

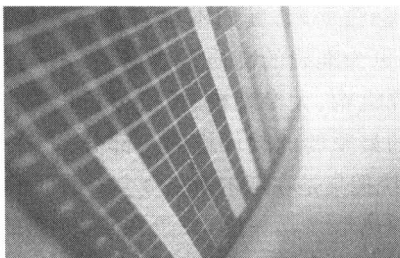
第2章 股票	(16)
2.1 标识	(16)
2.1.1 数据内容	(16)
2.1.2 数据查询及应用	(18)
2.2 事件	(20)
2.2.1 数据内容	(20)
2.2.2 数据展示及应用	(27)
2.3 行情与分配	(31)
2.3.1 数据内容	(31)
2.3.2 数据展示及应用	(34)
2.4 指数	(35)
2.4.1 数据内容	(35)
2.4.2 数据展示及应用	(36)
2.5 收益	(36)
2.5.1 数据内容	(36)
2.5.2 数据展示及应用	(38)
2.6 风险因子与三因子模型	(41)
2.6.1 数据内容	(41)
2.6.2 数据展示及应用	(41)
2.7 权证	(41)
2.7.1 数据内容	(41)
2.7.2 数据展示及应用	(43)
2.8 港股	(43)
2.8.1 数据内容	(43)
2.8.2 数据展示及应用	(43)
第3章 上市公司财务	(44)
3.1 财务报表	(44)
3.1.1 数据内容	(44)
3.1.2 数据查询及运用	(55)
3.2 财务比率	(55)
3.2.1 数据内容	(55)
3.2.2 数据查询及运用	(56)
3.3 股东权益和资产减值准备	(56)
3.3.1 数据内容	(56)
3.3.2 数据查询及运用	(57)
3.4 公司治理	(57)
3.4.1 数据内容	(57)
3.4.2 数据查询及运用	(59)

3.5 公司重大事项	(60)
3.5.1 数据内容	(60)
3.5.2 数据查询与使用	(62)
第4章 基金	(64)
4.1 标识与信息	(64)
4.1.1 数据内容	(64)
4.1.2 数据查询及应用	(66)
4.2 事件	(67)
4.2.1 数据内容	(67)
4.2.2 数据查询及运用	(70)
4.3 行情与净值	(70)
4.3.1 数据内容	(70)
4.3.2 数据展示及应用	(71)
4.4 投资组合	(71)
4.4.1 数据内容	(71)
4.4.2 数据展示及应用	(72)
4.5 财务指标	(72)
4.5.1 数据内容	(72)
4.5.2 数据展示及应用	(74)
第5章 固定收益	(76)
5.1 银行存款利率	(76)
5.1.1 数据内容	(76)
5.1.2 数据查询及应用	(77)
5.2 基准利率	(77)
5.2.1 数据内容	(77)
5.2.2 数据展现与应用	(80)
5.3 标识与信息	(80)
5.3.1 数据内容	(80)
5.3.2 数据展现与应用	(83)
5.4 事件	(86)
5.4.1 数据内容	(86)
5.4.2 数据展现与应用	(87)
5.5 交易所交易数据	(88)
5.5.1 数据内容	(88)
5.5.2 数据展现与应用	(90)
5.6 银行间交易数据	(91)
5.6.1 数据内容	(91)

5.6.2 数据展现与应用	(96)
5.7 债券指数	(98)
5.7.1 数据内容	(98)
5.7.2 数据展现与应用	(99)
第6章 宏观与行业	(101)
6.1 编码对照	(101)
6.1.1 数据内容	(101)
6.1.2 数据查询及应用	(102)
6.2 国民经济核算	(102)
6.2.1 数据内容	(102)
6.2.2 数据查询及应用	(103)
6.3 人口	(103)
6.3.1 数据内容	(103)
6.3.2 数据查询及应用	(104)
6.4 宏观金融统计	(104)
6.4.1 数据内容	(104)
6.4.2 数据查询及应用	(105)
6.5 主要产品产量和价格数据统计	(105)
6.5.1 数据内容	(105)
6.5.2 数据查询及应用	(106)
6.6 农业数据统计	(106)
6.6.1 数据内容	(106)
6.6.2 数据查询及应用	(107)
第7章 CRSP	(108)
7.1 日数据	(110)
7.1.1 数据内容	(110)
7.1.2 数据查询及应用	(110)
7.2 月数据	(114)
7.2.1 数据内容	(114)
7.2.2 累积收益与持有期收益之间的差别	(114)
7.2.3 没有交易时成交量却不为零的情况	(115)
7.3 事件数据	(116)
7.3.1 数据内容	(116)
7.3.2 数据查询及应用	(118)
7.4 最新文件	(119)
7.4.1 数据内容	(120)
7.5 CRSP 指数数据	(121)

7.5.1 数据内容	(121)
7.5.2 数据查询与应用	(123)
7.6 债券、国库券和通货膨胀	(124)
7.6.1 数据内容	(125)
7.6.2 数据查询及应用	(128)
7.7 CRSP 共同基金	(129)
7.7.1 数据内容	(129)
7.7.2 数据查询及应用	(133)
7.8 如何从 CRSP 获取数据	(134)
7.9 系统偏差和初步筛选	(136)
7.9.1 系统偏差	(136)
7.9.2 退市收益相关偏误	(137)
7.9.3 股票分配价格调整相关偏误	(138)
7.9.4 从日收益计算月收益	(138)
7.9.5 如何获取标准普尔 500(S&P500)数据集	(138)
7.9.6 数据筛选基础	(141)
7.9.7 处理缺失值(缺损值)	(143)
第 8 章 Compustat	(145)
8.1 行业年度数据	(148)
8.1.1 数据内容	(148)
8.1.2 数据查询及应用	(149)
8.2 分类数据	(164)
8.2.1 数据内容	(164)
8.2.2 数据查询及应用	(164)
8.3 管理层薪酬数据库	(166)
8.3.1 数据内容	(166)
8.3.2 数据查询及应用	(167)
8.4 Compustat 时点数据	(170)
8.4.1 数据内容	(170)
第 9 章 IBES	(175)
9.1 基本介绍和与 IBES 相关的研究主题	(175)
9.1.1 市场有效性理论 vs. 搜集公司信息的用处	(175)
9.1.2 利益冲突	(176)
9.1.3 声望影响	(176)
9.1.4 意见分散度是一种风险	(176)
9.1.5 意见分散度不是一种风险	(177)
9.1.6 公共信息 vs. 私人信息	(177)

9.1.7	IBES vs. TAQ 数据	(177)
9.1.8	预测准确性 vs. 好的推荐	(177)
9.1.9	基本介绍	(178)
9.2	详细历史数据	(178)
9.2.1	数据内容	(179)
9.2.2	数据查询及应用	(180)
9.3	概要历史数据	(183)
9.4	推荐数据集	(186)
9.5	停止数据集	(188)
9.6	除外文件	(189)
9.7	将 IBES 与 CRSP 或 Compustat 合并	(189)
9.8	其他	(190)
9.9	附录	(191)
第 10 章	NYSE TAQ (高频数据)	(193)
10.1	综合报价数据库集	(195)
10.2	综合交易数据库集	(199)
10.3	TAQ 名称数据集	(201)
10.4	主文件和红利文件	(202)
10.5	将 TAQ 和 CRSP 中的信息相匹配	(204)
10.6	如何得到区间交易数据	(204)
附录 1	RESET/DB 数据查询操作练习	(207)
附录 2	通过 ODBC 连接访问 RESET/DB	(231)
附录 3	RESET/DB 数据表名称对照表	(242)
附录 4	SAS 金融数据处理综合练习题	(251)



第 1 章

金融数据库概论

金融数据库,就是综合金融理论与计算机技术,将金融以及其他相关数据进行加工整理而成的,能够为金融教学、研究、金融投资等提供数据与相关服务的“数据平台”。请注意这里用的是数据平台。因此,这里定义的金融数据库不仅仅是数据库本身,它还包括基于数据库的相关数据处理、计算、建模及技术支持等服务。

国际著名金融数据库有芝加哥大学的证券价格研究中心(Center for Research in Security Prices, CRSP)和标准普尔公司(Standard and Poor's)的 Compustat 等。

国际著名数据库服务技术支持平台有美国宾夕法尼亚大学沃顿商学院研究数据服务中心(Wharton Research Data Services, WRDS);为机构投资者提供更专业服务的还有 Factset。

国内的 RESSET 锐思数据(www.resset.cn),是一个既提供国内经济、金融等数据,又提供相关服务的数据平台。

1.1 金融数据库的起源

实证金融与金融计量研究的发展,金融统计学、金融数学等学科的兴起,一个重要的基础是金融数据的搜集。对于学界的研究者,唯有搜集到全面、准确的数据,才能进行有意义的经济与金融研究。对于金融机构从业人员,金融数据库是他们从事投资研究的必要条件。

最初的研究者,对于其所需的金融数据,多数是通过自己搜集来完成。据统计,早期的实证研究者,在数据搜集整理上花费的时间要占总研究时间一半以上,甚至 would 超过 80%。

由于对数据的需求越来越迫切,因此提供统一、便捷的数据平台便成为必要。在很多研究机构,人们越来越注重数据的搜集、整理和共享。于是,产生了金融数据库的雏形。

然而,如果每个金融研究机构都进行数据的搜集、整理,就存在大量重复劳动,而且很多数据,普通研究机构难以搜集到,成本也很高。于是,专门提供金融数据库的公司出现了,金融数据库作为一个产品、一种服务,逐渐被人们所接受,这也意味着一个新兴行业的诞生。

金融数据库的出现,大大节省了研究者的时间。研究者不必再花费巨大的精力进行数据搜集和处理,而可以专注于研究本身。

此外,由第三方提供的金融数据库,其成本也较自己搜集更低,而且专门的金融数据库供应机构,能提供更为全面和更为准确的高质量数据。

金融数据库是各金融研究机构、大专院校、投资分析机构的重要数据平台。金融数据库是进行金融实证与投资研究的最重要的软环境的一部分。

股票市场价格的第一次系统搜集是在考尔斯基金会资助下于 20 世纪 30 年代编辑完成的。

考尔斯本人对股票市场的热情终其一生,只是在 1929 年的大灾难中稍受影响。如今隶属于耶鲁大学经济系的考尔斯基金会,是经济计量学在 20 世纪 40~50 年代大多数基础研究成就的源头。

今天,股票价格考尔斯指数早已经被更加详尽的计算机化数据库所取代,比如芝加哥大学证券价格研究中心(CRSP)数据库。反过来,这些数据库也有效地激励了金融领域实证研究及模型的检验,正是后者塑造了该领域的特殊风格。

目前,在国内外,很多金融数据库提供商都已经发展并成熟起来。

1.2 金融数据库的作用

金融数据库作为面向研究者和投资者提供数据支撑的基础平台,在金融实证和投资研究中,起着举足轻重的作用。

1. 节省研究者搜集数据的时间

前面提到,早期实证研究者在数据搜集整理上花费的时间占总研究时间一半以上,有时甚至会超过 80%。由此可以想象出,最初的金融研究人员在为某个特定研究专题搜集数据时,花费了怎样的精力和时间。所以,有了金融数据库,研究人员可以直接得到研究所需要的数据,无需自己花大量时间进行相关数据搜集工作。

2. 节省研究者整理数据的时间

金融数据库除了提供全面的基础数据外,还提供一些经过处理的数据,以满足研究者需要的数据合并、衍生指标等。数据整理的难度往往很大,技术含量亦很高。好的金融数据库会完成这些基础数据整理工作,研究者可以直接使用这些整理后的数据。

3. 降低研究者和研究机构的数据获得成本

直接购买专业金融研究数据库,会大幅降低研究者,特别是研究机构获取数据的成本。

4. 可以为研究者提供多种专业化的配套服务

专业金融数据库都有专业配套服务,包括特殊需求数据整理、计算程序设计等服务,支持研究,方便研究者使用。

5. 标准统一,有利于研究结果的比较

采用同一金融数据库有利于研究结果的比较。如果每个研究者都使用自己整理的数据库,相关指标计算标准不统一,其研究结果自然也不具有可比性。

1.3 金融数据库的类别

目前,金融数据库已经得到相当的发展。按照不同的分类标准,可将其划分为多种类型。

1. 按照内容和用途不同,可以将金融数据库分为行情咨询数据库和实证研究数据库两大类

- 行情咨询数据库

此类数据库主要是提供即时行情等咨询类数据。强调数据的实时性、直接性、准确性,而不需要对数据进行加工整理。此类数据库主要是为投资分析服务,其用户主要是各类投资者。早期的金融数据库多属此类。

- 实证研究数据库

顾名思义,实证研究数据库是为研究者进行实证研究服务。与行情咨询类数据不同,此类数据库通常为了便于各类研究者使用,包含更为全面的数据,强调数据的深度和广度;通常,此类数据库不仅提供原始数据,还为研究者提供一系列“衍生数据”,即对原始数据进行深层和专业的加工、整理,提供各类常用数据指标,以方便研究者使用,节省研究者时间。

随着技术发展,以及金融研究者和各类金融从业人员对金融数据需求范围不断扩大,很多金融数据库产品提供商都开始同时提供即时行情咨询类数据和实证研究数据,形成一个完整的数据平台。

2. 按照金融数据的种类不同,可以将金融数据库分为单一金融数据库和综合金融数据库

- 单一金融数据库

金融市场就其交易的金融产品不同,可以分为股票市场、债券市场、外汇市场、固定收益市场、金融期货市场等。相应的,单一金融数据库主要提供单一种类金融市场相关数据,如股票数据库、外汇数据库、债券数据库等。最初的金融数据库集中在股票数据上。这类单一金融数据库,数据范围相对狭窄,但是通常更为精专,例如可以提供更为多样的衍生指标和专题数据。

- 综合金融数据库

与单一金融数据库相对应,综合金融数据库通常提供多个金融市场相关数据。很多机构的投资范围通常跨了多个金融市场,同时研究者也希望进行这类跨市场的研究,这就需要多个市场的数据。越来越多的数据库提供商都在扩展数据覆盖范围,数据库综合化成为一种趋势。除了金融数据之外,很多研究往往需要涉及利率、通货膨胀等宏观经济数据,以及各类行业数据,分析师对各公司收益的预测,甚至法律信息,因此,很多金融数据库也把宏观数据、行业数据、法律信息等数据纳入了数据提供范围。金融数据库也成为名副其实的综合数据库。

1.4 中外金融数据库概况

国际著名的金融数据库都是由金融数据库专家经过多年研制发展与完善起来的。影响较大的有下列数据库:

- CRSP(证券价格研究中心,芝加哥大学)
- Compustat(美国公司及其他国家公司数据,标准普尔公司)
- NYSE TAQ(日内高频数据,纽约股票交易所)
- PACAP(Pacific-Basin Capital Markets,太平洋地区资本市场,美国罗德岛大学)

- Datastream
- IBES(Institutional Broker's Estimate System,机构经纪人估计系统)
- SDC(美国国家数据中心)
- GovPX(固定收益债券价格基准)
- Reuter(路透)
- Bloomberg(彭博)

国内的金融数据库起步较晚,有些数据库的设计者本身可能没有作过金融实证与投资研究,再加上一些研究人员不考查数据质量就作研究,所以,建立国内金融研究数据库的标准还需要一定时日。

根据本章后面提出的金融数据库选择标准,本书选择锐思数据(RESSET)作为中国金融研究数据库的样本。

1.4.1 CRSP

证券价格研究中心(CRSP)由芝加哥大学商学院研究生院于1959年成立。

1. 数据来源

- 美国股票和指数数据:纽约交易所和费城交易所数据
- 美国股票和指数数据:纳斯达克数据
- 美国国库券数据:多种来源
- 美国共同基金数据:多种来源

2. 数据范围

- 美国股票数据
- 美国股票指数
- 美国国库券
- 美国共同基金

3. 更新频率

- 年度更新:每年4月
- 季度更新:下一季度的第12个工作日
- 月度更新:下一月的第12个工作日

4. 四种CRSP数据

- 证券或指数的整体描述或总结数据
- 时间序列观测值数据
- 事件数据
- 日期数据

5. 两种数据库格式

- CRSP Access
- SFA(现已不用)

它们的差别是指数和组合数据的组织方式不一样。CRSP Access和SFA格式的股票和指数文件数据库结构是为了格式化和编程目的而设计的。

6. 提供数据文件格式

- C
- FORTRAN
- SAS
- ACCESS
- CRSP 提供的专用软件

7. 网址

<http://www.crsp.uchicago.edu>

1.4.2 Compustat

标准普尔 Compustat 提供北美地区大多数上市公司的年度和季度收益表、资产负债表、现金流量表以及其他附加数据。金融数据是通过一系列途径获得的,包括有线新闻网、新闻发布会、股东报告、与公司的直接联系,以及证券交易委员会公布的季度和年度文件。

Compustat 文件也包括关于合并、行业分割、银行、市场价格、股利发放以及收益等信息。

Compustat 提供自 1950 年到现在的相关数据集。

Compustat 包括的数据文件有:

- 行业文件(INA, INQ)
- 全覆盖文件(FCA, FCQ)
- 研究文件(RES, REQ)
- 银行文件(BNA, BNQ)
- 商业信息和行业分割(BIF)
- 价格、股利与收益文件(PDE)

文件说明:

行业文件——包括最近 20 年 NYSE 和 AMEX 上市公司的资产负债表、收益表和现金流量表。

全覆盖文件——包括在纳斯达克以及其他地区交易所上市公司的普通股交易记录以及全资子公司的优先股或债券交易记录。

研究文件——包含由于收购、合并、破产、清算、反收购、杠杆收购或者私有化等原因而没有包括在行业文件和全覆盖文件里面的公司信息。最新文件包含最近 20 年数据并且每年更新一次。有两个附加研究数据文件,一个包含 1961—1980 年的数据(每年更新一次),另外一个包含 1950—1969 年的数据(不会更新)。

银行文件——600~700 家银行数据。

商业信息和行业分割——包含每个公司 7 个财政年度数据。每年有 1~10 条关于行业分割的数据,这取决于公司报告。

需要一个公司信息时,可以搜索 Compustat 数据库(行业、研究和全覆盖文件)。WRDS 已经把这三个文件合成为 SAS 文件(COMPANN. SSD01 和 COMPQTR. SSD01)。

网址:

<http://www.kellogg.nwu.edu/researchcomputing/Compustat.htm>

1.4.3 PACAP

太平洋地区资本市场(The Pacific-Basin Capital Markets,PACAP)数据库是由罗德岛大学经济管理学院 Sandra Ann Morsilli 环太平洋资本研究中心开发的。PACAP 数据库包括太平洋地区 9 个国家和地区资本市场的历史数据。

表 1.1 国家(地区)及数据覆盖的年限

中国香港	1975—2003 年
日本	1975—2004 年
马来西亚	1975—2001 年
中国台湾	1975—2002 年
印度尼西亚	1975—2000 年
泰国	1975—2002 年
新加坡	1975—1998 年

表 1.2 可得到的文件

文件号	文件名称	文件号	文件名称
1	公司文件	7	月度市场收益文件
2	每日汇率文件	8	月度指数文件
3	主要经济统计文件——月度	9	日市场收益文件
4	主要经济统计文件——季度/年度	10	日指数文件
5	日期文件	11	月度股票价格和收益文件
6	金融状况文件——行业、公司	12	日股票价格和收益文件

目前,PACAP 文件包含从 1975 年 1 月 1 日至不同年份的数据(终止年份按国家或地区的不同而不同),记录是以国家为基础的,每年 9 月份进行年度更新。股票市场和公司信息是从各个国家或地区股票交易所搜集,经济统计资料来自其公共出版物和中央银行公布的数据。机器可以阅读的数据要进行内部一致性检验,原始数据用来检验有疑问的信息,非机器阅读信息是手写形式并经过核实。

网址:

<http://www.cba.uri.edu/PACAP>

1.4.4 Datastream

Datastream 提供了一个随时更新的包含全球经济、公司、市场等信息的数据平台入口。金融数据可以以文本和图形格式得到。

覆盖的数据类型:

股票,债券,公司账目,股票市场指数,利率和汇率,经济统计数据,期货和期权,抵押贷款以及商品。

数据来源:

股票市场指数来自于 WI Carr,IFC,JP Morgan 以及 Stanley;债券指数来自于 ISMA, Salomon Brothers 以及 Paribas;经济数据来自于德意志联邦银行、国际货币基金组织以及经合组织。